The Veterinary Journal 238 (2018) 15–21

Contents lists available at ScienceDirect

The Veterinary Journal

journal homepage: www.elsevier.com/locate/tvjl

Original Article

Bayesian latent class estimation of sensitivity and specificity parameters of diagnostic tests for bovine tuberculosis in chronically infected herds in Northern Ireland

A. Lahuerta-Marin[a,*], M.G. Milne[a], J. McNair[a], R.A. Skuce[a], S.H. McBride[a], F.D. Menzies[b], S.J.W. McDowell[a], A.W. Byrne[a,d], I.G. Handel[c], B.M. de C. Bronsvoort[c]

[a] Agri-Food and Biosciences Institute (AFBI), Veterinary Sciences Division, Stoney Road, Belfast BT4 3SD, UK
[b] Veterinary Epidemiology Unit, Department of Agriculture, Environment and Rural Affairs (DAERA), Dundonald House, Belfast, BT4 3SB, UK
[c] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Edinburgh EH25 9RG, UK
[d] School of Biological Sciences, Queen's University Belfast, Belfast BT7 1NN, UK

ARTICLE INFO

Article history:
Accepted 26 April 2018

Keywords:
Bovine tuberculosis
Bayesian latent class analysis
Interferon-γ test
Sensitivity
Specificity
Single intradermal cervical comparative tuberculin test

ABSTRACT

In the European Union, the recommended ante-mortem diagnostic methods for bovine tuberculosis (bTB) include the single intradermal cervical comparative tuberculin (SICCT) test and the interferon-gamma (IFN-γ) test as an ancillary test. The SICCT test has a moderate sensitivity (Se) and high specificity (Sp), while the IFN-γ test has good Se, but a lower Sp than the SICCT test. A retrospective Bayesian latent class analysis was conducted on 71,185 cattle from 806 herds chronically infected with bTB distributed across Northern Ireland (NI) to estimate the Se and Sp of the common ante-mortem tests and meat inspection. Analyses were also performed on data stratified by farming type and herd location to explore possible differences in test performance given the heterogeneity in the population. The mean estimates in chronically infected herds were: (1) 'standard' SICCT: Se 40.5–57.7%, Sp 96.3–99.7%; (2) 'severe' SICCT: Se 49.0%–60.6%, Sp 94.4–99.4%; (3) IFN-γ(bovine–avian) using a NI optical density (OD) cut-off difference of 0.05: IFN-$\gamma(B-A)_{NI}$: Se 85.8–93.0%, Sp 75.6–96.2%; (4) IFN-γ(bovine–avian) using a standard 'commercial' OD cut-off difference of 0.1: IFN-$\gamma(B-A)_{0.1}$: Se 83.1–92.1%, Sp 83.1–97.3%; and (5) meat inspection: Se 49.0–57.1% Se, Sp 99.1–100%. Se estimates were lower in cattle from dairy farms than from beef farms. There were no notable differences in estimates by location of herds. Certain population characteristics, such as production type, might influence the ability of bTB tests to disclose truly infected cases.

Introduction

Bovine tuberculosis (bTB) is a chronic infectious disease caused by *Mycobacterium bovis* and cattle are considered to be the main reservoir of infection (Allen et al., 2010). Bovine tuberculosis is a serious animal health problem that leads to economic and international trade restrictions for those countries that are not officially tuberculosis free (OTF) (Zinsstag et al., 2006; Good and Duignan, 2011). bTB has important economic consequences in Northern Ireland (NI), where cattle are an important part of the local economy and bTB is considered to be endemic.[1] Despite an ongoing eradication scheme, costing ~£30 million per annum,[2] bTB has proved difficult to eliminate, perhaps because of the complex and multifactorial nature of the disease (Humblet et al., 2009). Some herds in NI are also 'chronically infected', and experience prolonged breakdowns or recurrent bTB infection (Doyle et al., 2016).

The eradication programme in NI is based on a 'test and-cull' strategy. Two of the recommended ante-mortem screening tests for bTB include the single intradermal comparative cervical tuberculin (SICCT) test and the interferon gamma (IFN-γ) test, which is approved as an ancillary test as specified in European Union (EU) Council Directive EC/1226/2002 amending Annex B to Directive 64/432/EEC.[3] Both tests have well documented limitations in terms of performance characteristics (de la Rua-Domenech et al., 2006). The SICCT test can have a 'standard' or 'severe' interpretation. A standard interpretation is read when there is a difference of greater than 4 mm between the thickness at the site of injection of the bovine antigen and the site of injection of the avian antigen. A severe interpretation is one in which the thickness at the

* Corresponding author.
E-mail address: angela.lahuerta-marin@afbini.gov.uk (A. Lahuerta-Marin).
[1] See: https://www.daera-ni.gov.uk/publications/tuberculosis-disease-statistics-northern-ireland-2014 (accessed 8 October 2016).
[2] £1 = approx. US$1.41, €1.15 at 5 April 2018.
[3] See: https://publications.europa.eu/en/publication-detail/-/publication/b4b312f4-ba77-4cc7-99cd-5fc19c85d2cd/language-en (accessed 8 October 2016).

https://doi.org/10.1016/j.tvjl.2018.04.019

bovine site is greater than the avian site by 2 mm. The standard interpretation has a good specificity (Sp > 99%), but previous studies have suggested that it can have relatively poor sensitivity (Se 51–85%) (Nuñez-Garcia et al., 2017). This makes the SICCT test a good initial herd screening test, but with serious limitations if applied to chronically infected herds, because of the high probability of leaving undetected infection (Lahuerta-Marin et al., 2015, 2016).

Since 2004, the IFN-γ test has also been used in NI as a voluntary ancillary test in parallel with the SICCT test in high risk and chronically infected herds, with the main aim of identifying additional truly infected animals. Previous research indicates that the IFN-γ test using a standard 'commercial' optical density (OD) cut-off difference of 0.1 has a higher Se than the SICCT test (88–94%), but a lower Sp (85-98%) (de la Rua-Domenech et al., 2006; Clegg et al., 2011; Downs et al., 2011). In NI, a more stringent IFN-γ cut-off difference (OD 0.05) is applied than the commercial cut-off, which increases the Se of the test, but reduces the Sp. The application of both SICCT and IFN-γ tests together should improve the prospect of clearing bTB from herds where control has been problematic.

An estimate of the performance characteristics of each test is necessary in order to maximise the success in detecting infection, whilst minimising the number of false positive animals. This optimal scenario is difficult to achieve for bTB, where screening ante-mortem tests are imperfect, and there is no gold standard test, which makes it difficult to determine the true disease status of tested animals (Dohoo et al., 2009). To tackle the lack of a gold standard test, Bayesian latent class analysis has been used to estimate the performance characteristics of tests for bTB (Clegg et al., 2011; Alvarez et al., 2012; EFSA, 2012; Karolemeas et al., 2012; Bermingham et al., 2015; Nuñez-Garcia et al., 2017). However, previous attempts to derive estimates for Se and Sp in Northern Ireland using Bayesian latent class analysis (EFSA, 2012; Bermingham et al., 2015) have left some gaps in our understanding, namely: (1) the derivation of SICCT test parameters under both severe and standard interpretation; (2) the derivation of IFN-γ parameters under both standard commercial (cut-off OD difference 0.1) and NI (cut-off OD difference 0.05) interpretation; and (3) differences in test performance between beef and dairy cattle.

The aim of this study was to undertake new Bayesian latent class analysis to estimate the Se and Sp of each of the standard and severe SICCT tests, the IFN-γ(bovine–avian) test using a NI OD cut-off difference of 0.05, i.e. IFN-γ(B–A)$_{NI}$, the IFN-γ(B–A) using a standard 'commercial' OD cut-off difference of 0.1, i.e. IFN-γ(B–A)$_{0.1}$, the IFN-γ test using the 6 kDa early secretory antigenic target, i.e. IFN-γ(ESAT6), and detection of visible lesions consistent with bTB via meat inspection, on a robust dataset of animals from chronically infected herds tested across NI.

Materials and methods

Study population

The cohort of herds of cattle selected to participate in this retrospective study were based on their herd bTB histories. Only herds with chronic or recurrent infection, or very large recent breakdowns, were eligible for this cohort as part of the NI IFN-γ scheme from 2004 to 2010. The herd selection criteria for the IFN-γ scheme were applied as described by Lahuerta-Marin et al. (2015, 2016); ethical approval was not required for this study. Herd participation within the NI IFN-γ scheme was voluntary and the population under test was a convenience series; thus, the study population was not representative of the entire cattle population in NI. A total of 71,185 animals belonging to 806 cattle herds were included in the analysis; these animals were treated as a single population. The data set were curated and validated, and missing data were excluded. The analysis was carried out following the Standards for the Reporting of Diagnostic Accuracy Studies that Use Bayesian Latent Class Models (STARD-BLCM) guidelines (see Appendix: Supplementary material 1).[4]

Diagnostic testing

The SICCT test was performed according to requirements specified in EU Council Directive 64/432/EEC.[5] SICCT tests were performed by Department of Agriculture, Environment and Rural Affairs (DAERA) Veterinary Officers (VOs) and designated Local Veterinary Inspectors (LVIs) on behalf of DAERA within the participating Divisional Veterinary Office (DVO) regions. For IFN-γ testing, blood was collected during day 1 of the SICCT test before tuberculin injections by the DAERA VO, and samples were sent to AFBI for independent analysis within 8 h of collection. IFN-γ tests for purified protein derivatives (PPDs) for bovine–avian (B–A) and ESAT6 antigens were performed, as described by Welsh et al. (2002). For IFN-γ(A–B), two cut-off differences were used to interpret the results: (1) NI cut-off, with a net difference between the bovine PPD (PPDB) and avian PPD (PPDA) ODs (PPDB–PPDA) ≥ 0.05, if PPDB ≥ 0.1 OD; and (2) the commercially recommended cut-off, with a net difference PPDB–PPDA ≥ 0.1 OD. For IFN-γ(ESAT6), a net OD cut-off difference of 0.05 was used. Meat inspection was carried out as described in Lahuerta-Marin et al. (2016), whereby all animals sent to slaughter were assessed by meat inspectors for lesions consistent with bTB.

Bayesian latent class analysis

Hui and Walter (1980) published a Bayesian latent class model to evaluate diagnostic tests in the absence of a 'gold standard' test; some of the assumptions of the approach under this two-test, two-population latent class model are: (1) when multiple populations are being compared, each population prevalence should be different; (2) the Se and Sp of the test are the same across test populations; and (3) the tests are conditionally independent. Whilst the two-test, two-population approach is the 'classical' latent class model, the framework can be extended to any scenario whereby $S \geq R/(2^{R-1}-1)$, where S = number of populations and R = number of tests. Thus, a three-test, one population study system as presented here should be sufficient (Toft et al., 2005). The selection of a single study population, as opposed to two or more populations, reflects the limitations that come from identifying sub-populations which differ in prevalence, but not in how the each sub-population reacts to the test (Toft et al., 2005).

The analysis was performed as described by Hui and Walter (1980) and Bronsvoort et al. (2009), using three tests and one population. This model assumes that, for the population under study, the counts of the different combinations of test results for the three tests follow a multinomial distribution (i.e. +++; ++−; −++; −+−; +−+; +−−; −−+; −−−). In the model, seven parameters in total must be estimated, comprising the prevalence of the disease in the population and both the Se and Sp for each of the three tests. The data provide seven degrees of freedom; therefore the number of degrees of freedom are equal to the number of parameters to be estimated when conditional independence is assumed (Jones et al., 2010; Lewis and Torgerson, 2012). However, both the SICCT and IFN-γ tests are based on cell-mediated immunity and, therefore, are not independent. The implications of conditional dependence between tests were explored by running separate models accounting for model co-variance (see Appendix: Supplementary material 2: Part 1); however these models may be considered non-identifiable, requiring nine parameters to be estimated with only seven degrees of freedom (Toft et al., 2007; Jones et al., 2010). Where model selection took place, lower deviance information criterion (DIC) values were used to identify better fitting models.

Our model compared the distribution of results from various combinations of tests, i.e. SICCT standard, SICCT severe, IFN-γ(B–A)$_{NI}$, IFN-γ(B–A)$_{0.1}$, IFN-γ(ESAT6) and meat inspection, to derive estimates for the Se and Sp of each test (Toft et al., 2005). Where meat inspection data were available (n = 49,540), the first and second tests were the SICCT test and one of the IFN-γ tests, respectively, while the third test was the meat inspection data. However, parameter estimates for the whole population were derived using only SICCT, IFN-γ(B–A) and IFN-γ(ESAT6) tests, since meat inspection data were not available for the whole population. The latent state being estimated was the 'true' infection status of animals, which was defined as an animal infected with *M. bovis*; we made no inference as to whether the hosts were infectious, latently infected or had active infection.

All models were implemented in JAGS within the R statistical software environment, using packages *rjags* and *runjags*. Estimates of the prevalence were allowed to range from 0.1 to 0.3, following a uniform distribution. Parameter values for the Se and Sp of each test were estimated using flat, vague priors (*beta* (*1,1*)), although the impact of changing the priors was also explored (see Appendix: Supplementary material 2: Part 2). Three Monte–Carlo Markov chains (MCMCs) were each run for a total of 50,000 iterations, with the first 20,000 iterations discarded as a 'burn-in' and the subsequent 30,000 iterations retained for posterior inference, with 'thinning' performed every 10 iterations. The estimates and the Bayesian credibility intervals were reported; the credibility interval represents the limits in which the parameter estimate falls, with 95% credibility. The outputs, including the MCMC trace-plots, posterior density distribution plots and cross-correlation plots were assessed further to ensure that autocorrelation was low. Chain convergence after the initial burn-in was assessed by visual inspection of the MCMCs and via diagnostics proposed by Brooks and Gelman (1998), whereby

[4] See: http://www.equator-network.org/reporting-guidelines/stard-blcm/ (accessed 8 October 2016).

[5] See: https://www.daera-ni.gov.uk/articles/bovine-tuberculosis-tb-legislation (accessed 8 October 2016).